

Pubblicato su AgendaDigitale (NetworkDigital360) il 7 Giugno 2023 <https://www.agendadigitale.eu/cultura-digitale/nessun-dato-e-al-sicuro-dalla-reidentificazione-le-misure-che-servono-pe-la-nostra-privacy/> con il titolo “Nessun dato è al sicuro dalla “reidentificazione”: le misure che servono per la nostra privacy”

E' anonimo se così pare

Avv. Silvio Noce - Avvocato esperto di privacy, diritto delle nuove tecnologie, amministrazione digitale, collabora con numerose realtà pubbliche tra cui Regione Emilia-Romagna, Lepida Scpa e Regione Toscana.

Dott. Marco Manca - Presidente di SCImPULSE Foundation (NL) e coordinatore del programma di liaison Europeo della Società Italiana di Telemedicina.

Premessa

Lo scorso 26 aprile il Tribunale dell'Unione europea ha pronunciato una controversa sentenza nella causa T-557/20, con le parti Single Resolution Board (SRB) contro European Data Protection Supervisor (EDPS).

Nell'ambito di una procedura di risoluzione, di cui al regolamento europeo n. 806/2014, cui è stato sottoposto il Banco Popular Espanol, con relativa vendita dell'attività di impresa, SRB, al fine di definire la necessità di riconoscere in capo agli azionisti un indennizzo, ha invitato questi a manifestare il loro interesse ad esercitare il diritto di essere ascoltati ai sensi dell'articolo 41, paragrafo 2, lettera a), della Carta dei diritti fondamentali dell'Unione europea. Il procedimento di “ascolto” consta di due fasi, la prima di iscrizione e la seconda di consultazione. In tale seconda fase, gli azionisti e i creditori interessati hanno presentato osservazioni sulla decisione preliminare nonché sulla versione non riservata della “valutazione 3”. In tale contesto, SRB ha chiesto a Deloitte, nella sua qualità di valutatore indipendente, *“di valutare le osservazioni pertinenti relative alla valutazione 3, di fornirgli un documento contenente la sua valutazione e di esaminare se la valutazione 3 restasse valida alla luce di tali osservazioni”*. Per far ciò SRB ha trasmesso a Deloitte le osservazioni ricevute nella fase di consultazione recanti un codice alfanumerico. Mediante tale codice, solo SRB avrebbe potuto collegare le osservazioni ai dati ricevuti durante la fase di iscrizione. A seguito di una serie di reclami, il garante europeo della protezione dei dati (“GEPD”) ha contestato a SRB la violazione dell'articolo 15 del regolamento 2018/1725 in quanto non aveva informato i reclamanti, nell'informativa sulla protezione dei dati personali, che era possibile che i loro dati personali fossero comunicati a Deloitte.

Di seguito i punti di attenzione rilevati.

Il ruolo di Deloitte

Nella versione in lingua italiana della sentenza succitata è facilmente riscontrabile un errore di "traduzione". Al punto 32 n. 2 della stessa sentenza la versione in lingua inglese¹ riporta che *"The fact that Deloitte was not mentioned in SRB's [privacy statement] as a potential recipient of the personal data collected and processed by the SRB **as the controller** in the context of the [right to be heard] process constitutes an infringement of the information obligations laid down in Article 15(1)(d) [of Regulation 2018/1725]."*. Ovvero viene rappresentato dai Giudici (e tale elemento non è oggetto di contestazione delle parti) che Deloitte assume il ruolo di titolare del trattamento, mentre nella versione in lingua italiana del pronunciamento è riportato che *"Il fatto che Deloitte non sia stata menzionata nella informativa sulla protezione dei dati personali del CRU quale potenziale destinatario dei dati personali raccolti e trattati dal CRU, nella sua qualità di **responsabile del trattamento** nell'ambito della procedura relativa al diritto di essere ascoltato, costituisce una violazione dell'obbligo di informazione previsto all'articolo 15, paragrafo 1, lettera d), [del regolamento 2018/1725]"*.

Tale aspetto è dirimente.

Giova preliminarmente sottolineare che l'art. 5 par. 1 n. 9² del GDPR definisce destinatario dei dati come colui che riceve comunicazione di dati personali, che si tratti o meno di terzi. All'art. 2ter del Codice per la protezione dei dati personali viene, altresì, specificato che per comunicazione deve intendersi dare conoscenza dei dati personali a uno o più soggetti determinati diversi dall'interessato, dal responsabile del trattamento o da persone autorizzate.

Come ci ricorda l'EDPB nelle "Linee guida 07/2020 sui concetti di titolare del trattamento e di responsabile del trattamento ai sensi del GDPR"³, "destinatario" è concetto relazionale, nel senso che rimanda a una relazione con un titolare o con un responsabile del trattamento da una prospettiva specifica. Nella fattispecie considerata la prospettiva è costituita dall'essere Deloitte organismo indipendente ai sensi dell'art. 20 del regolamento n. 806/2014 e, quindi, legittimata a ricevere i dati.

Nel caso in cui Deloitte fosse stata considerata responsabile del trattamento, la vexata questio in ordine all'anonimizzazione effettiva dei dati alla stessa trasmessi non si sarebbe posta, in ragione del fatto che i dati sarebbero circolati nel medesimo perimetro di titolarità di SRB e l'utilizzo di codici identificativi unici al posto dei dati identificativi diretti, sarebbe stata considerata misura di sicurezza correlata ai profili di autorizzazione concessi.

¹ <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62020TJ0557>

² Coincidente con definizione dell'articolo 3, punto 13, del regolamento 2018/1725

³ https://edpb.europa.eu/system/files/2022-02/eppb_guidelines_202007_controllerprocessor_final_it.pdf

Pseudonimizzazione alla luce del “tempo” e stato dell’arte

In punto di re-identificazione e/o riconducibilità di un codice pseudonimo e delle informazioni ad esso correlate alla categoria di dati personali, la sentenza fa un salto temporale indietro di almeno 25 anni, poichè accenna in maniera molto approssimativa ai concetti di “mezzi legali” che consentono la de-identificazione, del punto di vista del destinatario del dato e lo fa imperniando le proprie argomentazioni sui connotati essenziali della sentenza del 19 ottobre 2016, Breyer (C-582/14, EU:C:2016:779) che concerne ambiti di trattamenti di dati personali, ovvero mole e tipologia di dati profondamente diversi.

Per spiegare tale perentorietà nel giudizio, è necessario partire dal presupposto che, come già osservato *“De-identification can reduce the privacy risk associated with collecting, processing, archiving, distributing or publishing information. De-identification thus attempts to balance the contradictory goals of using and sharing personal information while protecting privacy”*⁴. Ovvero le tecniche di anonimizzazione e de-identificazione (che in tale contesto utilizziamo come sinonimi) non consentono di eliminare il “privacy risk”⁵, sulla base di due ulteriori elementi chiave da considerare, ovvero il tempo e un nuovo senso al concetto di “stato dell’arte”.

Per quel che concerne il primo, in prima battuta deve essere osservato che non può essere accolto il principio per cui il “mittente” dei dati debba definire la disponibilità -giuridica, materiale, potenziale!- di dati “ausiliari” da parte del destinatario dei dati. Anche perchè, come già sostenuto, *“Va sottolineato come un fattore in grado di compromettere significativamente le tutele introdotte con l’anonimizzazione sia la disponibilità di dati ausiliari riferibili ad una persona a cui collegare il dato anonimizzato.*

Poiché la quantità di informazioni, anche pubblicamente disponibili, è destinata a crescere nel tempo, un mezzo oggi valutato irragionevole, in considerazione dell’informazione ausiliaria attualmente disponibile, potrà non essere giudicato tale in successive valutazioni, anche tenuto conto dell’evoluzione delle tecnologie. Pertanto, la considerazione sui mezzi non deve essere vista come una valutazione una tantum, ma come un’operazione che deve essere oggetto di un riesame periodico in ragione dei nuovi rischi connessi alla crescente disponibilità di mezzi tecnici a basso costo (il cloud

⁴ NISTIR 8053, De-Identification of Personal Information, Simon L. Garfinkel Information Access Division Information Technology Laboratory, October 2015

⁵ Ancora il NIST *“ISO/TS 25237:2008(E) provides explanatory text stating: “NOTE—Anonymization is another subcategory of de-identification. Unlike pseudonymization, it does not provide a means by which the information may be linked to the same person across multiple data records or information systems. Hence reidentification of anonymized data is not possible.” [p. 6] The problem with these definitions is that some anonymization attempts have resulted in data have been re-identified, implying that the data thought to be anonymized actually weren’t”.*

*computing ad esempio), all'accessibilità pubblica sempre maggiore di altre banche dati e alle competenze tecniche utilizzabili"*⁶.

Il GDPR ci impone di considerare, nell'analisi dell'adeguatezza delle misure di sicurezza da implementare, lo *"stato dell'arte e dei costi di attuazione rispetto ai rischi che presentano i trattamenti e alla natura dei dati personali da proteggere"*.

Ebbene, un approccio improntato alla tutela sostanziale dei diritti e delle libertà degli interessati, proprio in virtù del criterio dello *"stato dell'arte"*, ci impone di considerare che la mole di dati è esponenzialmente crescente e, quindi, i mezzi che oggi sono *"irragionevolmente utilizzabili"* fra qualche ora potrebbe non essere più cotanto irragionevoli; d'altra parte, se considerare lo stato dell'arte significa analizzare e tenere conto dei trend in materia di cybersecurity, non può, oggi, essere trascurata una riflessione sulla *"tecnica" Harvest Now, Decrypt Later*".

Da ciò non deve derivare un atteggiamento di rassegnazione; è necessario sviluppare tecniche di anonimizzazione poiché *"quanto minore sarà il numero di elementi potenzialmente identificativi presenti nel dato anonimizzato, tanto maggiore sarà lo sforzo necessario all'identificazione della persona per chi utilizza quel dato"*⁷; l'intento è rendere poco appetibile, in ragione dello sforzo richiesto, il nostro set di dati.

D'altra parte, le tecniche di anonimizzazione e di pseudonimizzazione hanno l'obiettivo, rispettivamente, di eliminare e limitare l'associabilità tra un set di dati e gli interessati. *"Questo tipo di protezione è in genere finalizzato a contrastare le azioni compiute da un attaccante per eseguire una reidentificazione"*⁸. L'attaccante può essere interno ed esterno. In tale secondo caso dovranno certamente essere considerati gli elementi succitati relativi a disponibilità esponenziale di dati ulteriori e capacità computazionale.

D'altra parte, le conseguenze che potrebbero derivare dai risvolti applicativi dei principi definiti nella succitata sentenza minerebbero profondamente l'efficacia e l'efficienza dei modelli di elaborazione dei dati. In particolare, in tema di *"opinioni personali"*, in una coorte molto poco significativa (come nel caso di specie), l'indagine in ordine all'applicabilità delle cautele della normativa in materia di protezione dei dati personali dovrà coinvolgere ogni singola opinione, posto che ciascuna di queste può riportare stati e informazioni che consentano la riconducibilità alla persona⁹. Tale modello è certamente

⁶ Big data e Privacy by design, di G. D'Acquisto

⁷ ancora G. D'Acquisto

⁸ ENISA, TECNICHE DI PSEUDONIMIZZAZIONE E MIGLIORI PRATICHE, Raccomandazioni per sviluppare tecnologie conformi alle disposizioni in materia di protezione dei dati e privacy, NOVEMBRE 2019

⁹ Nel caso di specie, se uno degli interessati avesse risposto *"Io, che ho guidato la BCE tra il 2000 e il 2003, so che la scelta di vendere l'attività d'impresa ci causerà danni economici certi, e quindi dovrà essere riconosciuto in capo agli azionisti un indennizzo cospicuo"*.

incompatibile con lo stato dell'arte di elaborazione dei dati, in cui ad una crescente capacità computazionale deve corrispondere un efficiente paradigma di composizione del dato.

Considerazioni tecniche

Raggruppare caratteristiche identificative come "età" e "titolo di lavoro" in intervalli, piuttosto che includere le cifre esatte, o tokenizzarle, rimuoverle perfino, può aiutare a offuscare i dati a prima vista. Tanti sono gli stratagemmi proposti (per esempio lo spostamento delle date di un periodo di tempo casuale), ma i metodi devono essere scelti con cura per non oscurare le relazioni per le quali i dati erano interessanti in primo luogo. È comunemente accettato che esista un trade-off tra il grado di anonimizzazione e l'utilità finale dei dati. Un'altra regola empirica è che maggiore sia il numero di attributi in un set di dati, maggiore sarà la probabilità che una corrispondenza sia corretta e quindi minore la probabilità che i dati possano essere protetti dall'"anonimizzazione". Tuttavia, è tutt'altro che accettato che rendere anonimi i dati sia possibile tout court.

Già nel 2019 Rocher, L., Hendrickx, J.M. & de Montjoye, YA pubblicavano su *Nature Communication* ¹⁰ il loro metodo di reidentificazione di datasets anonimi, che otteneva un risultato del 99,98% di corretta reidentificazione degli individui in set di dati anonimi con soli 15 attributi "demografici".

Già da anni la ricerca suggerisce quanto sia facile reidentificare gli individui all'interno di set di dati anonimi (per esempio ¹¹ e ¹²) ma l'articolo citato propone addirittura un modello statistico per stimare quanto facile sia la reidentificazione per un set di dati arbitrario calcolando la probabilità che una potenziale corrispondenza sia corretta (altresì detto, valutando quanto "uniche" siano le corrispondenze dei data point in popolazioni di dati che, come quelli umani, hanno strutture -sparsità, boundary conditions, correlazioni interne, ...) ed i risultati mostrano che pochi attributi sono spesso sufficienti per reidentificare con elevata fiducia gli individui anche in set di dati fortemente incompleti. In un articolo del 2021¹³ inoltre, viene dimostrato che i rischi di reidentificazione rimangono elevati anche nei set di dati su scala nazionale, quando si possa inferire informazione circa la geolocalizzazione, ed è teoricamente possibile

¹⁰ Rocher, L., Hendrickx, J.M. & de Montjoye, YA. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun* **10**, 3069 (2019). <https://doi.org/10.1038/s41467-019-10933-3>

¹¹ de Montjoye, YA., Hidalgo, C., Verleysen, M. *et al.* Unique in the Crowd: The privacy bounds of human mobility. *Sci Rep* **3**, 1376 (2013). <https://doi.org/10.1038/srep01376>

¹² Yves-Alexandre de Montjoye et al., Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science* **347**, 536-539 (2015). DOI:10.1126/science.1256297

¹³ Ali Farzanehfar, Florimond Houssiau, Yves-Alexandre de Montjoye. The risk of re-identification remains high even in country-scale location datasets, *Patterns*, 2021;2(3):100204. DOI: 10.1016/j.patter.2021.100204.

identificare il 93% delle persone in un set di dati contenente informazioni su 60 milioni di individui, utilizzando solo quattro dimensioni/informazioni ausiliarie.

Si respinge così l'affermazione che il campionamento o il rilascio di set di dati parziali (ad esempio, da una rete ospedaliera o da un singolo servizio online) forniscano una negabilità plausibile. Inoltre, anche accettando l'argomento a volte proposto che l'univocità di alcune popolazioni sia bassa (un argomento spesso utilizzato per giustificare che i dati sono sufficientemente anonimizzati per essere considerati anonimi), molti individui restano a rischio di essere reidentificati con successo da un utente malintenzionato semplicemente sfruttando i metodi dimostrati negli articoli citati.

Questo suggerisce che nessun set di big data "anonimizzato" e rilasciato possa essere considerato al sicuro dalla reidentificazione, non senza rigorosi controlli di accesso i cui costi e meccanismi di fall-back e mitigazione del rischio di intrusione, andrebbero valutati attentamente caso per caso, e documentati.

Va inoltre considerato che la ricerca in profilazione e reidentificazione è estremamente attiva ed avanzata, e Facebook (oggi Meta, ndr) già nel 2017-2018 si trovò al centro di un'indagine per acclarare come potesse incrociare dati "anonimi" raccolti dai navigatori del web, ivi inclusi coloro che non avessero affatto dei profili Facebook, con i dati di contatto degli utenti registrati, per ricostruire le informazioni sensibili necessarie alla compagnia ad offrire il proprio servizio di marketing profilato^{14 15 16}.

Un ultimo commento lo merita il modello di rischio apparentemente selezionato dalla Corte:

1. se è possibile (sospendiamo il giudizio su questo, ma notiamo che non esiste un registro trasparente pubblico delle aziende e dei loro accessi ai dati) verificare che un ricevente non abbia accesso, legittimamente, ai dati necessari a reidentificare un dato set di dati nel presente, NON pare però che esista neppure vincolo contrattuale alcuno ad impedirne l'acquisizione a posteriori;
2. non è in essere alcun meccanismo di policy che tenga traccia degli incroci possibili tra dataset i responsabili dei dati, e non è in essere alcun iter per la notifica ai responsabili ed a cascata ai soggetti;
3. non è solo dal rischio derivante dagli accessi legittimi ai dati con potenziale di deanonimizzazione che il GDPR intende tutelare il cittadino.

In sostanza, esistono oggi due pesi e due misure per quanto riguarda la sicurezza delle nostre infrastrutture e dei nostri cittadini online: da una parte la sentenza del 26 Aprile

¹⁴ <https://about.fb.com/news/2018/04/data-off-facebook/>

¹⁵ <https://www.theverge.com/2018/4/11/17225482/facebook-shadow-profiles-zuckerberg-congress-data-privacy>

¹⁶ <https://gizmodo.com/how-facebook-figures-out-everyone-youve-ever-met-1819822691>

sembra invocare un pragmatismo mercato centrico; dall'altra le misure di irrobustimento delle infrastrutture per il remoto (ancora oggi) rischio computazione quantistica promuovono l'adozione di nuovi standard di crittazione per difendersi dagli scenario HNDL (Harvest Now, Decrypt Later -> Raccogli adesso, decrittata più tardi) stabilendo un principio di accountability, cui la compliance deve allinearsi. Nello stesso modo in cui i nostri dati vanno tutelati tramite crittazioni quantum-ready sin da oggi, pena il rischio che nel prossimo futuro (qualora il quantum computing realizzasse le sue promesse) dati sensibili scambiati oggi in modo sicuro divenissero improvvisamente disponibili ad organizzazioni criminali o persino stati in conflitto¹⁷ (carte di credito scambiate con protocolli di crittazione sicuri oggi, ma non robusti ad un attacco da computer quantistico, ad esempio), motivo per cui iniziative come quella del NIST¹⁸ producono raccomandazioni sin da oggi... non dovremmo attenderci lo stesso livello di difesa anche per i dati (gli stessi spesso, ndr) che vengono resi disponibili per usi secondari?

Le autorità di regolamentazione ed i legislatori dovrebbero insomma riconoscere la minaccia rappresentata dalla reidentificazione dei dati e prestare attenzione legale a "sistemi e misure di sicurezza dimostrabili per il miglioramento della privacy" tra cui quelli citati nel documento del 2015¹⁹ dove si discutono metodi come la ricerca crittografata e la privacy preserving computation; meccanismi di controllo granulare degli accessi e tracciamento dell'origine dei dati (e qui si apre un capitolo che meriterà un altro articolo, sul ruolo del contesto e sulla auspicabilità di integrare considerazioni legate alla verifica di appropriatezza ed autorizzazione nei sistemi di identità digitale wallet che sono in corso di pilot attualmente in EU); e strumenti di certificazione delle policies di accesso associate ad ogni istanza a supporto di pratiche di auditing ed analisi di near miss.

Confidiamo, pertanto, in un secondo grado di giudizio maggiormente consapevole

¹⁷<https://www.whitehouse.gov/briefing-room/statements-releases/2022/05/04/national-security-memorandum-on-promoting-united-states-leadership-in-quantum-computing-while-mitigating-risks-to-vulnerable-cryptographic-systems/>

¹⁸<https://www.nist.gov/news-events/news/2022/07/nist-announces-first-four-quantum-resistant-cryptographic-algorithms>

¹⁹ Giuseppe D'Acquisto, Josep Domingo-Ferrer, Panayiotis Kikiras, Vicenç Torra, Yves-Alexandre de Montjoye, Athena Bourka. Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics. <https://arxiv.org/abs/1512.06000>